①

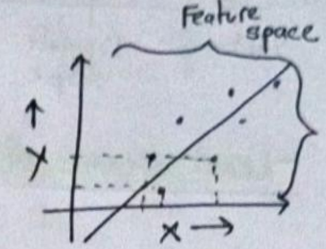
INTRODUCTION TO LINEAR REGRESSION

- Regression is supervised learning.

Given − X → Y
(Input)

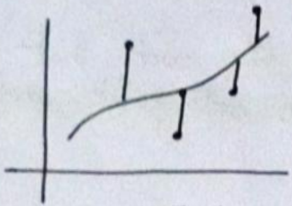
y is continuous

X → Predict Y.
(unseen)



- X can be comprised of one or more features.
- In linear regression, it will be a straight line. (Find the best fit line)

How to find best fit line? → Find the distance between line and point (each point).



→ Use a loss function that measure the square error in prediction of y (x) from x.

→ Sum of square error is a famous method.

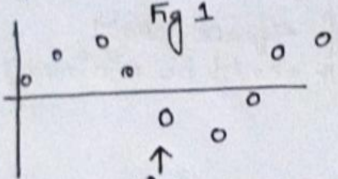


Fig 1
Very high SSE

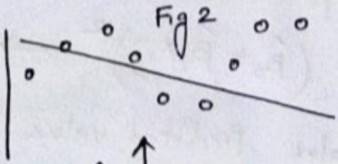


Fig 2
High SSE

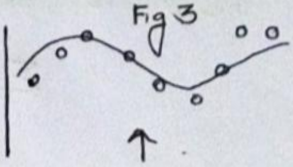


Fig 3
Minimum SSE

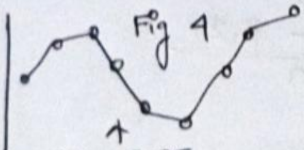
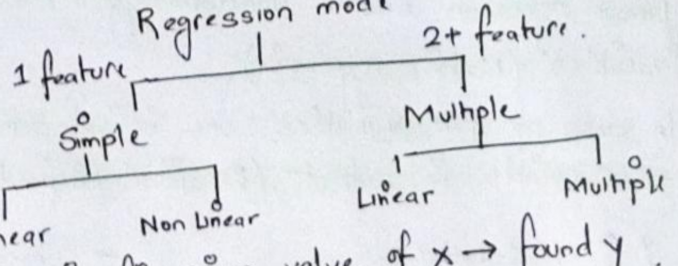


Fig 4
Zero SSE

→ Fig 4 have zero SSE (choose the lowest SSE) but it is not generalised (overfit the data).

→ So choose Fig 3.

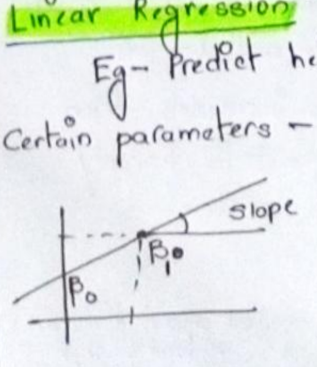
Type of regression model −



(Definition)
Linear Regression → A straight line fn. given value of x → found y.

Eg− Predict height from age, predict house price from house area.

Certain parameters − ① Intercept (Intercept y axis) ⓘⓘ Slope.



$\boxed{Y = \beta_0 + \beta_1 X}$.

Intercept slope .

There may be some noise in data (ϵ) epsilion. (error)

$\boxed{Y = \beta_0 + \beta_1 x + \epsilon}$

eg − i) Based on cost of sensor, we want to measure how much distance it cover, error is quality (+, -) of the sensor.

$$\boxed{y = \beta_0 + \beta_1 x + \varepsilon}$$

Y intercept   Populate slope   Random error.

- Multiple linear regression.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2^2 + \beta_3 x_3^3 + \dots + \beta_p x_p^1 + \varepsilon.$$

In this we have p variable (no of variables)

- Least square line - Least square regression line is the line that makes the vertical distance from the data points to the regression line as small as possible. It is called least square because the best line of fit is one that minimize the variance (sum of square of the errors).

- Assumption about error → Expected (Error) = 0.
  → Error are independent.
  → Error are normally distributed. Mean = 0.

This ~~to~~ error known as gaussian noise (white noise).

- The least square regression line is the unique ~~[illegible]~~ line such that the sum of squared vertical (y) distance between data points and the line is smallest possible.

$$\Sigma \left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_1) \right)^2 \longrightarrow \text{Sum of Square error (It should be minimum)}$$

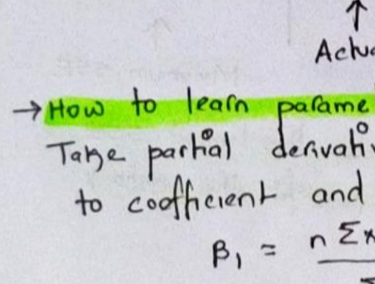
Actual value   Predicted value

→ How to learn parameter ? β_0 & β_1.
Take partial derivatives of the objective function (SSE) with respect to coefficient and set these to 0. and solve.

$$\beta_1 = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2} \qquad \beta_0 = \frac{\Sigma y - \beta_1 \Sigma x}{n}$$

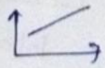
→ Linear regression aids in understanding the relationship between two or more variables (Multiple regression).

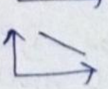
→ In case of two variables, one is independent Variable (Input) and the other variable is output (Predicted / dependent Variable).

→ 

b_0 → Y intercept
b_1 → Slope.

- Suppose price of house, Independent variable will be square of house in sqm². and dependent variable will be price.
- So, if area of house increases then price also increases so it is linearly relationship between the variables

→ When x↑ y↑, Slope is +ve ($\hat{y} = b_0 + b_1 x$). (Intercept) 

→ When x↑ y↓, Slope is -ve ($\hat{y} = b_0 - b_1 x$) (Intercept) 

student spend time on social networks, then grade of subject / result↓

③

For example,

(INPUT) x	(OUTPUT) y	$x-\bar{x}$	$(x-\bar{x})^2$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$\hat{y}$	$(\hat{y}-y)$	$(\hat{y}-y)^2$
1	2	-2	4	-2	4	2.8	0.8	0.64
2	4	-1	1	0	0	3.4	0.6	0.36
3	5	0	0	1	0	4	-1	1
4	4	1	1	0	0	4.6	0.6	0.36
5	5	2	4	1	2	5.2	0.2	0.4
			10		6	↑		

$\bar{x}$ → Mean of x, $\bar{y}$ → Mean of y $\bar{x}=3$, $\bar{y}=4$ Refer to eqn (i)

$$b_1 = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2} = \frac{6}{10} = 0.6$$

$\hat{y} = b_0 + b_1 x$

$4 = b_0 + 0.6(3)$

b_0 is calculated using mean coordinated $(3,4)$.

$$\hat{y} = b_0 + b_1 x$$
$$4 = b_0 + (0.6)(3)$$
$$b_0 = 4 - (0.6)(3)$$
$$b_0 = 2.2$$

So, $\hat{y} = 2.2 + 0.6x$ (Regression line) — (i)

Evaluate the model,

Standard error of estimate $= \sqrt{\dfrac{\Sigma(\hat{y}-y)^2}{n-2}}$

↓

Calculate the difference between actual and predicted.

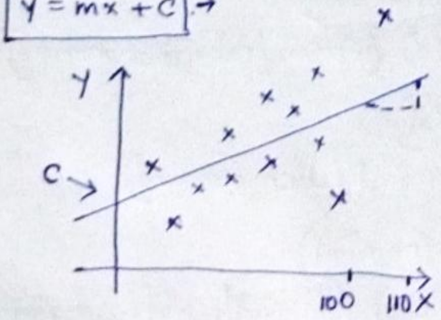
In case of 1 row.

$\sqrt{\dfrac{2.4}{5-2}} = 0.89$.

In most cases, if error is less than 1 then it is acceptable.

—x—

$\boxed{y = mx + c}$ →



$y = mx + c$

when $x = 0$, we can find value of c that is constant or intercept.

$y = m(0) + c$

$y = c$.

for value of m, if we change a unit of x what is the change in y.

$100 \rightarrow y_1$ $110 \rightarrow y_2$.

$m = y_2 - y_1$

Linear → Linear means line.

What is linear relationship?

→ A linear relationship (or linear association) used to describe a straight line relationship between a variable and a constant, which follow $y=mx+c$.

Use of linear relationship?.

→ It is a correlation, which describe how one variable changes in linear relation or fashion to changes in another variable.

Example of Linear relationship → i) Calculate distance travelled by rate of speed over period of time.
ii) Price of house given the area of the house.

- Regression analysis can result in linear or non linear graphs.

→ **Linear regression** → Relationship between variables can be described with a straight line.

→ **Non linear regression** → It cannot be defined by straight line.

Types of Linear Regression —

1) **Simple linear regression** – 1 dependent variable (interval or ratio)
1 independent variable (interval or ratio or dichotomous).

2) **Multiple linear regression** – 1 dependent variable (interval or ratio)
2+ independent variables (interval or ratio or dichotomous).

3) **Logistic regression** - 1 dependent variable (dichotomous)
2+ independent variable(s) (interval/ratio/dichotomous)

4) **Ordinal regression** - 1 dependent variable (ordinal)
1+ independent variable (nominal/dichotomous)

5) **Multinomial regression** - 1 dependent variable (nominal)
1+ independent variable(s) (interval/ratio/dichotomous)

6) **Discriminat analysis** - 1 dependent variable (nominal)
1+ independent variable(s) (interval/ratio).

→ **Dichotomy** is a simply a split of a set into two mutually exclusive subset whose union is ~~a~~ the original set.

→ **Nominal data** are normally nouns, with no order in them. Eg - Country, gender, etc.
Ordinal data comes with level of orders. Eg - First, Second, Third etc.